AL/HR-TP-1997-0052

UNITED STATES AIR FORCE ARMSTRONG LABORATORY

A TALE OF TWO TEST BATTERIES: A COMPARISON OF THE AIR FORCE OFFICER QUALIFYING TEST AND THE MULTIDIMENSIONAL APTITUDE BATTERY

Thomas R. Carretta

HUMAN RESOURCES DIRECTORATE
Aircrew Training Research Division
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

Paul D. Retzlaff
Joseph D. Callister, Captain, USAF

AEROSPACE MEDICINE DIRECTORATE
Clinical Sciences Division
2507 Kennedy Circle
Brooks Air Force Base TX 78235-5117

Raymond E. King, Major, USAF

CREW SYSTEMS DIRECTORATE
Human Engineering Division
22 H Street
Wright-Patterson Air Force Base OH 45433-7022

December 1997

AIR FORCE MATERIEL COMMAND
ARMSTRONG LABORATORY
HUMAN RESOURCES DIRECTORATE
AIRCREW TRAINING RESEARCH DIVISION
6001 South Power Road, Building 558
Mesa AZ 85206-0904

NOTICE

THOMAS R. CARRETTA
Project Scientist

DEE H. ANDREWS
Technical Director

LYNN A. CARROLL, Colonel, USAF
Chief, Aircrew Training Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1997	Interim Paper - June 1997 to August 1997

4. TITLE AND SUBTITLE

A Tale of Two Test Batteries: A Comparison of the Air Force Officer Qualifying Test and the Multidimensional Aptitude Battery

5. FUNDING NUMBERS

PE - 62205F
PR - 1123
TA - A1
WU - 01

6. AUTHOR(S)

Carretta, T.R., Retzlaff, Paul D., Callister, Joseph D., King, Raymond E.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Armstrong Laboratory

Human Resources Directorate	Aerospace Medicine Dir	Crew Systems Dir
Aircrew Training Rsch Div	Clinical Sciences Div	Human Engineering Div
7909 Lindbergh Drive	2507 Kennedy Circle	2255 H Street
Brooks AFB TX 78235-5352	BAFB TX 78235-5117	WPAFB OH 45433-7022

8. PERFORMING ORGANIZATION

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory Human Resources Directorate
Aircrew Training Research Division
6001 S. Power Road, Bldg 561
Mesa AZ 85206-0904

10. SPONSORING/MONITORING

AL/HR-TP-1997-0052

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr Thomas R. Carretta, (510)536-3922

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Air Force Officer Qualifying Test (AFOQT) and the Multidimensional Aptitude Battery (MAB) were administered to 2,233 US Air Force pilot candidates to investigate the common sources of variance in these batteries. The AFOQT was operationally administered as part of the officer commissioning and aircrew selection testing requirement. The MAB is a clinical test battery and was administered to provide an intellectual baseline to assist clinicians when it becomes necessary to evaluate pilots with cognitive referral questions. A joint factor analysis of the AFOQT and MAB revealed that each battery had a hierarchical structure. The higher-order factor in the AFOQT previously had been identified as general cognitive ability (g). The intercorrelation between the higher-order factors from the batteries was .981, indicating that both measured g. Although both batteries measured g and included verbal, spatial, and perceptual speed tests, the AFOQT also included tests of aviation knowledge not found in the MAB. Additional studies are required to evaluate the utility of the AFOQT for clinical assessment and the MAB for officer and aircrew selection.

14. SUBJECT TERMS
AFOQT, Air Force Officer Qualifying Test, Aircrew selection, Cognitive ability, Comparative factor structure, Confirmatory factor structure, MAB, Multidimensional Aptitude Battery, Personnel measurement, Personnel selection

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

CONTENTS

FIGURES

TABLES

Preceding Page Blank

PREFACE

This effort was performed under Work Unit 1123A101, Pilot Selection and Classification Support, in support of Aircrew Selection and Classification (R&D). The authors thank Charles E. Lance, Malcolm James Ree, and William C. Tirre for their helpful comments.

Send correspondence and requests for reprints to the first author at AL/HRAA, 7909 Lindbergh Drive, Brooks Air Force Base, TX 78235-5352. Send electronic mail to carretta@alhrm.brooks.af.mil.

A TALE OF TWO TEST BATTERIES: A COMPARISON OF THE AIR FORCE OFFICER QUALIFYING TEST AND THE MULTIDIMENSIONAL APTITUDE BATTERY

SUMMARY

The Air Force Officer Qualifying Test (AFOQT) and Multidimensional Aptitude Battery (MAB) were administered to 2,233 US Air Force pilot candidates to investigate the common sources of variance in those batteries. The AFOQT was operationally administered as part of the officer commissioning and aircrew selection testing requirement. The MAB is a clinical test battery and was administered to provide an intellectual baseline to assist clinicians when it becomes necessary to evaluate pilots with cognitive referral questions. A joint factor analysis of the AFOQT and MAB revealed that each battery had an hierarchical structure. The higher-order factor in the AFOQT previously had been identified as general cognitive ability (*g*). The intercorrelation between the higher-order factors from the batteries was .981, indicating that both measured *g*. Although both batteries measured *g* and included verbal, spatial, and perceptual speed tests, the AFOQT also included tests of aviation knowledge not found in the MAB. Additional studies are required to evaluate the utility of the AFOQT for clinical assessment and the MAB for officer and aircrew selection.

INTRODUCTION

The Air Force Officer Qualifying Test (AFOQT) is used to qualify civilians and prior-enlisted US Air Force (USAF) personnel for officer commissioning through the Officer Training School and Reserve Officer Training Corps programs. It is also used to qualify applicants who pass other educational and physical requirements for aircrew training. The AFOQT has been validated for pilot and navigator training (Arth, Steuck, Sorrentino, & Burke, 1990; Carretta, 1992; Carretta & Ree, 1995; Koonce, 1982; Olea & Ree, 1994; Ree & Carretta, 1996; Ree, Carretta, & Teachout, 1995) and for several other officer jobs (Arth, 1986; Arth & Skinner, 1986; Finegold & Rogers, 1985).

In 1994, the Air Force Medical Operations Agency began a program to establish a psychological testing baseline for Air Force pilots. This baseline was intended to assist clinicians when evaluating pilots with cognitive referral questions (Callister, King, & Retzlaff, 1996; Retzlaff, Callister, & King, 1996). One of the tests used to establish this baseline is the Multidimensional Aptitude Battery (MAB) (Jackson, 1985). The MAB is normally administered in paper-and-pencil form. The USAF developed a computerized version which was administered to pilot candidates during a flight screening program (King & Flynn, 1995).

The purpose of this study was to determine the extent to which the AFOQT and MAB measure the same constructs. If there is considerable overlap between the two batteries, further research may be directed toward using the AFOQT for clinical assessment and the MAB for officer and aircrew selection.

METHOD

Participants

Participants were 2,233 US Air Force pilot candidates who completed the AFOQT and a computerized version of the MAB. The sample had a mean age of 20.6 years and was predominantly male (92%) and White (87%).

Measures

Air Force Officer Qualifying Test. The AFOQT is a paper-and-pencil multiple aptitude battery used for officer commissioning and aircrew training selection (Skinner & Ree, 1987). It is developed and maintained by the USAF. Administration time is about 4 hours. The 16 AFOQT tests are combined to create five operational composites: Verbal, Quantitative, Academic Aptitude, Pilot, and Navigator-Technical. It has an hierarchical factor structure and measures general cognitive ability (*g*) and the lower-order factors of verbal, math, spatial, aircrew interest/aptitude, and perceptual speed (Carretta & Ree, 1996).

Multidimensional Aptitude Battery. The MAB is a broad-based test of intellectual ability. It was patterned after the Wechsler Adult Intelligence Scale (WAIS-R; full-scale $\underline{r} = .91$). Although the MAB requires about the same amount of time to administer as the WAIS-R (about 1.5 hours), it can be group-administered and machine scored, while the WAIS-R cannot.

The paper-and-pencil version of the MAB was developed by Jackson (1985) and the computerized version by the USAF Armstrong Laboratory (Retzlaff, King, & Callister, 1995). The computerized version was developed and used with the consent of the test author with explicit copyright permission. The two versions have the same 10 tests with identical items. The tests are Information, Comprehension, Arithmetic, Similarities, Vocabulary, Digit Symbol, Picture Completion, Spatial, Picture Arrangement, and Object Assembly. These tests are combined to form three composites: Full Scale (all 10 tests), Verbal (first five tests), and Performance (last five tests).

The MAB was administered on a 386-based computer with a 14-inch color monitor. Participants entered their responses using a keypad and mouse or light pen.

Procedures

The AFOQT was completed as a requirement of application for officer commissioning and/or aircrew selection. The time frame for AFOQT-testing varied. Some took the AFOQT near the completion of high school or while in college. Others took it after completing college. All participants completed the MAB shortly before beginning the Enhanced Flight Screening Program. MAB testing was done to establish an ideographic cognitive baseline for the clinical evaluation of pilots for comparative purposes after sustaining a head injury or other neurological insult.

Analyses

The participants represented a range-restricted sample because they had already been selected for college and for an officer commissioning program based on AFOQT and/or college entrance exams. The Lawley correction procedure (Lawley, 1943; Ree, Carretta, Earles, & Albert, 1994) was applied to estimate the means, variances, and correlations of the tests as they would be found in USAF officer applicants (Skinner & Ree, 1987). The confirmatory factor analyses were conducted using the range-restriction-corrected data as it provided a superior estimate of the means, standard deviations, and correlations.

Hierarchical confirmatory factor analyses (HCFAs) were performed using LISREL 8 (Jöreskog & Sörbom, 1996). The first-order confirmatory factor analysis (CFA) allowed all observed variables (16 AFOQT and 10 MAB tests) to load on their first-order factors and those first-order factors to correlate with each other. The first-order factors included the five lower-order AFOQT factors of verbal, math spatial, aircrew interest/aptitude, and perceptual speed and two MAB factors representing the MAB Verbal (first five tests) and Performance (last five tests) composites. A higher-order CFA was then conducted using the first-order factor intercorrelation matrix. This higher-order CFA allowed the five AFOQT factors to load on a higher-order general factor (g_{AFOQT}) and the two MAB factors to load on a second higher-order general factor (g_{MAB}). These two general factors were allowed to correlate and between-battery relationships among the lower-order factors were examined. Generalized least squares estimation procedures were used.

Although it may appear that the higher-order g_{MAB} factor is underdefined with only two indicators, Costner (1969) discusses the circumstances under which two indicators are sufficient. Generally, it is not required that all correlations between different pairs of indicators be identical. Rather, it is required that several estimates of a single abstract coefficient (e.g., factor loading) be consistent.

Several fit indices were computed. These included the χ^2, Comparative Fit Index (CFI) (Bentler, 1990), Non-Normed Fit Index (NNFI) (Marsh, Balla, & McDonald, 1988), and Root Mean Square Error of Approximation (RMSEA) (Browne & Cudeck, 1993).

RESULTS AND DISCUSSION

Table 1 shows the means and standard deviations of the tests in observed and corrected-for-range-restriction form. The observed AFOQT means were on average about .90 standard deviations above the normative values and the variances were about 77 % of the normative values for USAF officer applicants (Skinner & Ree, 1987). The observed means for the MAB tests were about 1 standard deviation above the normative value of 50 and the variances were about 54% of the normative value of 100 for adults (Jackson, 1985). After correction for range restriction (to USAF officer applicant norms), the MAB tests were still about .62 standard deviations above their normative value and the variances were about 69% of the adult normative value of 100. This suggests that USAF officer applicants are above adult norms on the construct measured by the MAB (i.e., intellectual ability).

Table 1.
Means and Standard Deviations for AFOQT and MAB Scores

Score	Abbr.	Observed Mean	Observed SD	Corrected Mean	Corrected SD
AFOQT					
Verbal Analogies	VA	18.29	3.31	13.36	4.23
Arithmetic Reasoning	AR	18.43	4.57	11.00	4.40
Reading Comprehension	RC	17.93	4.34	15.83	5.93
Data Interpretation	DI	18.81	3.83	11.15	3.93
Word Knowledge	WK	16.86	4.84	13.28	5.83
Math Knowledge	MK	19.87	4.39	14.48	6.04
Mechanical Comp.	MC	11.60	3.72	9.78	3.65
Electrical Maze	EM	8.89	3.31	7.68	4.22
Scale Reading	SR	27.93	5.88	20.07	6.73
Instrument Comp.	IC	15.08	4.13	8.82	4.76
Block Counting	BC	14.22	3.44	10.62	4.39
Table Reading	TR	30.69	5.96	26.46	7.35
Aviation Information	AI	13.31	4.24	8.65	4.08
Rotated Blocks	RB	9.94	2.76	7.59	3.36
General Science	GS	11.43	3.52	8.54	3.66
Hidden Figures	HF	10.89	2.75	9.60	2.76
MAB					
Information	INF	66.80	6.89	64.36	7.18
Comprehension	COM	59.74	4.36	58.17	4.60
Arithmetic	ARI	60.89	6.23	54.72	6.60
Similarities	SIM	59.82	8.66	56.14	9.15
Vocabulary	VOC	60.29	9.33	58.15	10.02
Digit Symbol	DIG	63.10	6.98	58.15	7.81
Picture Completion	PC	59.47	6.43	56.44	6.79
Spatial	SPA	59.10	8.94	54.04	9.68
Picture Arrangement	PA	51.95	7.01	48.33	7.45
Object Assembly	OBJ	58.94	7.58	53.68	8.31

Note. Means and standard deviations were corrected for range restriction using the multivariate Lawley (1943) procedure. An AFOQT officer applicant sample was used (Skinner & Ree, 1987).

The correlations among the tests are shown in Table 2. The observed correlations (above the diagonal) were positive with two exceptions involving the AFOQT Aviation Information test and two MAB tests (AI and DIG = -.010; AI and SPA = -.007). The largest observed correlation was between two AFOQT math tests, AR and DI (.636).

Table 2.
Correlation Matrix for AFOQT and MAB Scores

Score	VA	AR	RC	DI	WK	MK	MC	EM	SR	IC	BC	TR	AI	RB	GS	HF	INF	COM	ARI	SIM	VOC	DIG	PC	SPA	PA	OBJ
VA	100	479	573	472	587	407	389	153	297	191	156	143	243	236	480	205	290	293	299	315	351	230	308	193	215	286
AR	580	100	445	636	438	620	379	220	558	156	273	270	228	287	456	240	222	255	558	233	210	284	177	187	200	285
RC	730	580	100	483	641	369	362	131	296	149	120	177	276	201	494	157	325	393	322	335	397	214	264	113	229	207
DI	530	670	550	100	445	471	333	200	512	184	251	296	294	276	404	234	227	249	452	247	217	290	195	162	220	258
WK	680	460	770	460	100	375	349	074	280	126	110	131	343	193	533	150	324	322	277	316	494	153	247	087	188	209
MK	550	710	510	600	400	100	283	131	414	072	190	201	113	233	456	222	234	217	427	227	211	301	162	167	170	274
MC	480	510	460	460	400	480	100	333	252	352	227	111	360	430	538	254	191	253	234	166	207	079	348	245	187	303
EM	270	370	230	380	170	400	440	100	259	302	344	243	119	311	199	264	082	027	173	073	022	169	209	253	150	244
SR	480	660	450	620	370	600	480	450	100	207	375	428	251	253	295	225	118	135	431	140	061	301	119	172	187	227
IC	340	410	330	430	280	390	490	440	490	100	346	266	299	330	214	277	065	084	089	074	032	100	248	235	189	217
BC	450	530	400	510	320	490	500	470	610	490	100	410	105	325	118	318	028	051	229	049	011	261	169	289	199	292
TR	340	440	350	470	270	440	300	310	560	340	510	100	134	175	075	199	015	030	209	072	010	318	077	123	165	154
AI	300	310	340	340	320	250	500	290	330	560	310	210	100	130	385	032	075	114	076	037	085	-010	095	-007	011	032
RB	430	470	350	420	290	490	540	420	490	460	550	340	340	100	298	352	120	111	209	111	084	197	321	404	210	368
GS	510	490	550	440	510	520	570	340	410	410	370	250	460	400	100	196	319	276	268	244	298	112	327	144	153	270
HF	400	400	360	390	310	400	390	340	470	360	450	360	270	420	340	100	124	087	173	118	083	156	256	263	171	298
INF	382	295	425	287	406	307	250	159	223	172	182	129	140	195	363	211	100	253	174	191	293	123	235	113	159	152
COM	401	335	495	309	424	296	304	088	238	186	211	145	171	191	320	190	316	100	233	369	302	134	225	069	177	148
ARI	418	608	436	518	340	534	363	303	528	283	424	350	162	358	331	305	242	300	100	198	187	287	124	163	157	215
SIM	427	324	455	322	425	320	242	148	258	177	209	191	110	200	304	217	263	432	277	100	301	232	254	106	197	267
VOC	450	270	516	264	587	254	242	071	160	114	156	120	117	144	321	174	355	375	246	377	100	159	187	077	158	123
DIG	415	449	385	445	292	493	274	304	469	265	442	459	117	373	259	302	219	233	412	325	243	100	214	255	286	314
PC	434	323	399	321	360	331	420	295	312	340	349	212	242	408	414	337	303	297	246	328	257	337	100	269	293	389
SPA	355	360	269	325	210	378	376	361	382	345	452	254	176	519	272	338	188	151	298	189	141	381	354	100	254	407
PA	364	343	371	350	318	334	301	240	348	289	351	283	140	319	261	267	236	259	272	280	238	391	372	339	100	376
OBJ	440	437	366	404	319	461	426	362	430	349	477	307	199	491	377	387	231	231	344	341	195	445	468	500	457	100

Note. Decimals were omitted to conserve space. Correlations above the diagonal were observed. Correlations below the diagonal were corrected for range restriction. Lawley's(1943) multivariate correction was applied to the tests. An AFOQT officer applicant sample was used as a reference group (Skinner & Ree, 1987).

All correlations were positive after correction for range restriction (below the diagonal). See Ree et al. (1994) for an explanation of change in correlation sign after correction for range restriction. The largest correlation after correction for range restriction was between two AFOQT verbal tests, RC and WK (.770) and the smallest correlation (.071) was between a spatial test from the AFOQT (EM) and a verbal test from the MAB (VOC).

The correlations among the 26 tests were used to estimate a seven-factor, first-order CFA (5 lower-order AFOQT factors and 2 lower-order MAB factors). The χ^2 (275) was 2,032.791, CFI was .974, the NNFI was .970, and the RMSEA was .053. This is evidence of a good fit. The factor loadings for this lower-order model are shown in Table A1. The resulting correlation matrix for the lower-order factors (Table 3) was used to estimate the hierarchical model.

Table 3 shows the correlations among the first-order factors. They ranged from .450 (aviation and MAB verbal) to .895 (AFOQT verbal and math) with a mean value of .727. An examination of the between-battery correlations showed the AFOQT verbal and math factors to have higher correlations with the MAB verbal factor, while the AFOQT spatial, aviation, and perceptual speed factors had higher correlations with the MAB performance factor. The MAB verbal factor showed its highest between-battery correlation with the AFOQT verbal factor (.893) and its lowest correlation with aviation (.450). The MAB performance factor had its highest between-battery correlation with spatial (.854) and its lowest correlation with aviation (.587). The correlation between the two MAB factors was .787.

Table 3.
First-Order Factor Intercorrelations

Factor[a]	Verbal	Math	Spatial	Aviation	Percep. Speed	MAB Verbal	MAB Performance
Verbal	1.000						
Math	0.895	1.000					
Spatial	0.781	0.825	1.000				
Aviation	0.560	0.652	0.808	1.000			
Perceptual Speed	0.651	0.719	0.834	0.677	1.000		
MAB Verbal	0.893	0.858	0.719	0.450	0.530	1.000	
MAB Performance	0.768	0.754	0.854	0.587	0.683	0.787	1.000

[a]The first five factors were from the AFOQT and the last two factors were from the MAB.

The hierarchical model is shown in Figure 1. The loadings of the lower-order factors on their respective higher-order factors were high, ranging from .775 to .976. This indicated that the lower-order factors were essentially measures of their respective higher-order factors. The strong correlation between the two higher-order factors (.981) indicated that they measured the same higher-order factor. Because of the strength of this correlation and because the higher-order AFOQT factor is known to be psychometric *g*, it is apparent that the higher-order factor in the MAB also is *g*. General cognitive ability accounted for more variance than the sum of the lower-order factors for both batteries. The proportion of common variance accounted for by *g* was similar for the two batteries: 67.2% for the AFOQT (Carretta & Ree, 1996) and 67.7% for the MAB.

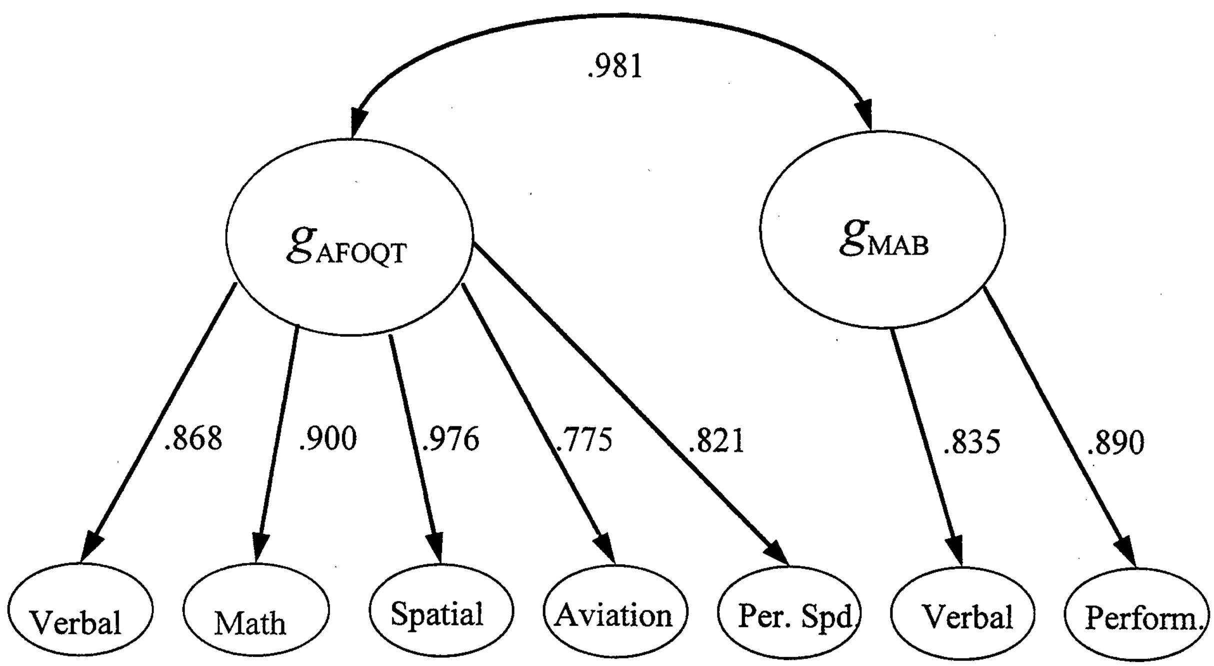

Figure 1. Hierarchical Model.
Note. The higher-order factors were g_{AFOQT} and g_{MAB}, respectively. The lower-order AFOQT factors were Verbal, Math, Spatial, Aviation Interest/Aptitude, and Perceptual Speed. The lower-order MAB factors were MAB Verbal and MAB Performance.

Similar results were reported by Sperl, Ree, and Steuck (1992) and by Stauffer, Ree, and Carretta (1996). Sperl et al. examined the relationship between the verbal and math tests from the AFOQT and Armed Services Vocational Aptitude Battery (ASVAB). They found a first canonical correlation between the two batteries of .93 indicating a high level of common variance. Stauffer et al. examined the common sources of variance between all 10 ASVAB tests and a set of computer-based cognitive components tests. As in the current study, Stauffer et al. found a strong correlation (.994) between the higher-order factors from the two batteries indicating both higher-order factors measured the same construct.

These results suggest that both the AFOQT and MAB may be acceptable for establishing a clinical cognitive baseline for USAF pilot trainees. Both batteries measure psychometric *g* as well as verbal, spatial, and perceptual speed (the later two factors are subsumed in the MAB performance factor). However, it is not clear that the two batteries identically measure the lower-order factors.

The chief advantage of the MAB over the AFOQT for use as a clinical assessment tool is its similarity to standard clinical intelligence tests such as the WAIS-R. Air Force clinical psychologists routinely use the WAIS-R to evaluate pilots referred for cognitive assessment. Because of its similarity to the WAIS-R, clinicians find it relatively easy to make pre- and post-incident comparisons using baseline MAB data. If the AFOQT were to be used instead of the MAB for making pre- and post-incident comparisons, clinicians would need training to become more familiar with the AFOQT and its relation to the WAIS-R or MAB.

Although the AFOQT takes longer to administer than the MAB (4 hours vs. 1.5 hours), it is already in operational use for officer commissioning and aircrew selection so would not require any special administration as does the MAB. Further, the AFOQT includes tests of aviation interest/aptitude not covered by the MAB (i.e., Instrument Comprehension and Aviation Information). These tests have been shown to be useful for predicting pilot performance beyond measures of *g* and specific cognitive abilities such as verbal, math, spatial, and perceptual speed (Olea & Ree, 1994; Ree & Carretta, 1996; Ree, Carretta, & Teachout, 1995). Therefore, if the MAB were to be used in place of the AFOQT, it would be desirable to retain at least the aviation interest/aptitude portions of the AFOQT to ensure no loss of validity for predicting pilot training performance.

Additional studies are planned to evaluate the utility of the AFOQT for clinical assessment and the utility of the MAB for officer and aircrew selection. If the two batteries are interchangeable, the Air Force may be able to save administration time by using one test for both purposes.

REFERENCES

Arth, T. O. (1986). *Validation of the AFOQT for non-rated officers* (AFHRL-TP-85-50, AD A164 134). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Arth, T. O., & Skinner, M. J. (1986). *Aptitude selection for Air Force officer non-aircrew jobs*. Paper presented at the 28th annual meeting of the Military Testing Association, Mystic, CT.

Arth, T. O., Steuck, K. W., Sorrentino, C. T., & Burke, E. F. (1990). *Air Force Officer Qualifying Test (AFOQT): Predictors of undergraduate pilot training and undergraduate navigator training* (AFHRL-TP-89-52). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-240.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Lang (Eds.). *Testing statistical equation models* (pp. 136-162). Newbury Park, CA: Sage.

Callister, J. D., King, R. E., & Retzlaff, P. D. (1996). Cognitive assessment of USAF pilot training candidates. *Aviation, Space, and Environmental Medicine, 67*, 1124-1129.

Carretta, T. R. (1992). Recent developments in U. S. Air Force pilot candidate selection and classification. *Aviation, Space, and Environmental Medicine*, *63*, 1112-1114.

Carretta, T. R., & Ree, M. J. (1995). Air Force Officer Qualifying Test Validity for predicting pilot training performance. *Journal of Business and Psychology*, *9*, 379-388.

Carretta, T. R., & Ree, M. J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology*, *8*, 29-42.

Costner, H. L. (1969). Theory, deduction, and rules of correspondence. *American Journal of Sociology*, *75*, 245-263.

Finegold, L., & Rogers, D. (1985). *Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training* (AFHRL-TR-85-13, AD A158 162). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Jackson, D. N. (1985). *Multidimensional aptitude battery*. Port Huron, MI: Research Psychologists Press.

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

King, R. E., & Flynn, C. F. (1995). Defining and measuring the "right stuff": Neuropsychiatrically enhanced flight screening (N-EFS). *Aviation, Space, and Environmental Medicine, 66*, 951-956.

Koonce, J. M. (1982).Validation of a proposed pilot trainee selection system. *Aviation, Space, and Environmental Medicine, 53*, 1166-1169.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62, Part 1*, 28-30.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79*, 845-849.

Ree, M. J., & Carretta, T. R. (1996). Central role of *g* in military pilot selection. *The International Journal of Aviation Psychology, 6*, 111-123.

Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology, 79*, 298-301.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*, 721-730.

Retzlaff, P. D., Callister, J. D., & King, R. E. (1996). *The computerized neuropsychological evaluation of US Air Force pilots: Clinical procedures and data-based decision* (AL/AO-TR-1996-0107). Brooks AFB, TX: Clinical Sciences Division, Armstrong Laboratory.

Retzlaff, P. D., King, R. E., & Callister, J. D. (1995). *Comparison of a computerized version to a paper/pencil version of the Multidimensional Aptitude Battery* (AL/AO-TR-1995-0121). Brooks AFB, TX: Clinical Sciences Division, Armstrong Laboratory.

Skinner, J., & Ree, M. J. (1987). *Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of form O* (AFHRL-TR-86-68). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Sperl, T. C., Ree, M. J., & Steuck, K. W. (1992). Armed Services Vocational Aptitude Battery and Air Force Officer Qualifying Test: Analyses of common attributes. *Military Psychology, 4*, 175-188.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *The Journal of General Psychology, 123*, 193-205.

APPENDIX A:

Confirmatory Factor Analysis Solution for the Seven-Factor First-Order Model

Table A1.
Factor Loadings for the Seven-Factor Lower-Order Model

Score	Verbal	Math	Spatial	Aviation	Percep. Speed	MAB Verbal	MAB Performance
VA	0.838						
AR		0.845					
RC	0.896						
DI		0.767					
WK	0.864						
MK		0.795					
MC			0.781				
EM			0.547				
SR		0.386			0.471		
IC				0.794			
BC			0.454		0.321		
TR					0.666		
AI				0.756			
RB			0.702				
GS	0.515			0.322			
HF			0.570				
INF						0.524	
COM						0.596	
ARI						0.662	
SIM						0.597	
VOC						0.649	
DIG							0.648
PC							0.652
SPA							0.597
PA							0.580
OBJ							0.715